BIG DATA HADOOP AND SPARK DEVELOPMENT

ASSIGNMENT – 14

Table of Contents:

BIG DATA HADOOP AND SPARK DEVELOPMENT

1. Introduction

In this assignment, the given task is performed and Output of the task is performed and Screenshots are attached.

2. Objective

This assignment consolidates the deeper understanding of the Session – 14 Introduction to Scala, which stands for Scalable Language.

3. Problem Statement

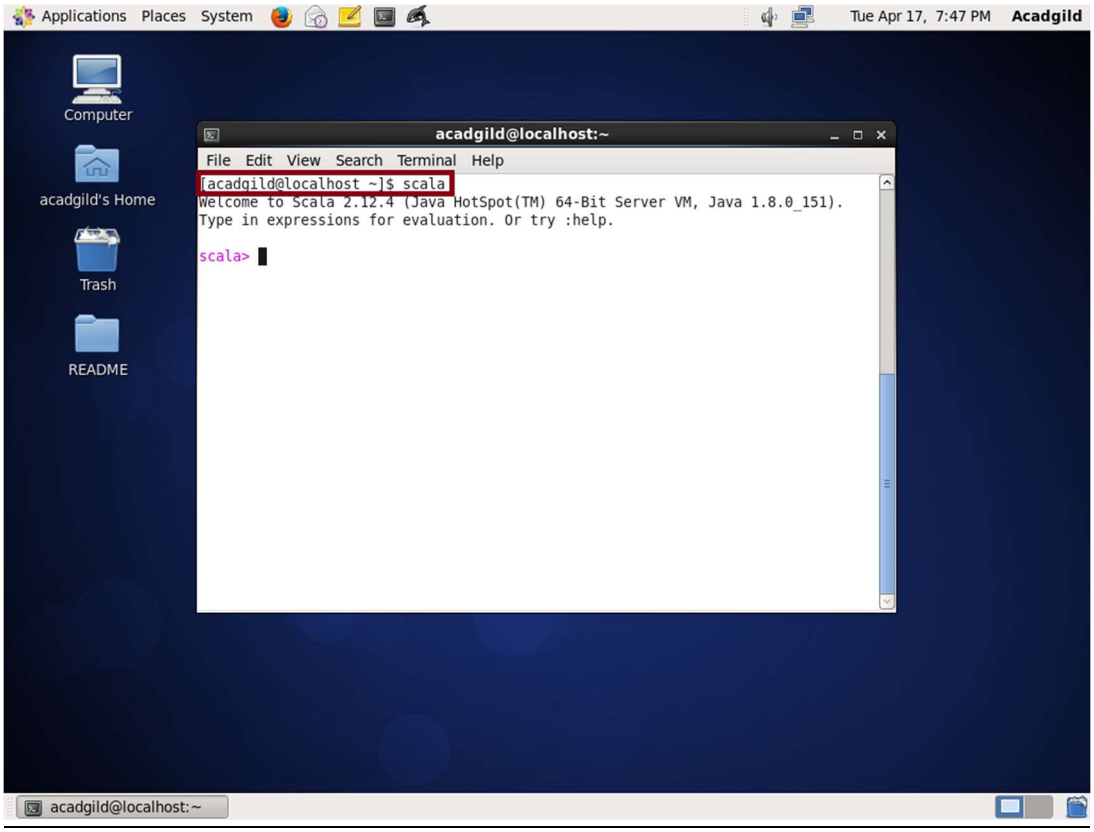
- **Task 1**
  1. Given a list of strings - List[String] (“alpha”, “gamma”, “omega”, “zeta”, “beta”) Find count of all strings with length 4.
  2. Convert the list of string to a list of integers, where each string is mapped to its corresponding length.
  3. Find count of all strings which contain alphabet ‘m’. - Find the count of all strings which start with the alphabet ‘a’.

- **Task 2**
  1. Create a list of tuples, where the 1st element of the tuple is an int and the second element is a string. Example - ((1, ‘alpha’), (2, ‘beta’), (3, ‘gamma’), (4, ‘zeta’), (5, ‘omega’))
  2. For the above list, print the numbers where the corresponding string length is 4.
  3. find the average of all numbers, where the corresponding string contains alphabet ‘m’ or alphabet ‘z’.

4. Expected Output

Preparing to perform tasks:

Start Scalable language console by **scala** command.



❖ Task 1

- **Given a list of strings - List[String] (“alpha”, “gamma”, “omega”, “zeta”, “beta”) - Find count of all strings with length 4**.

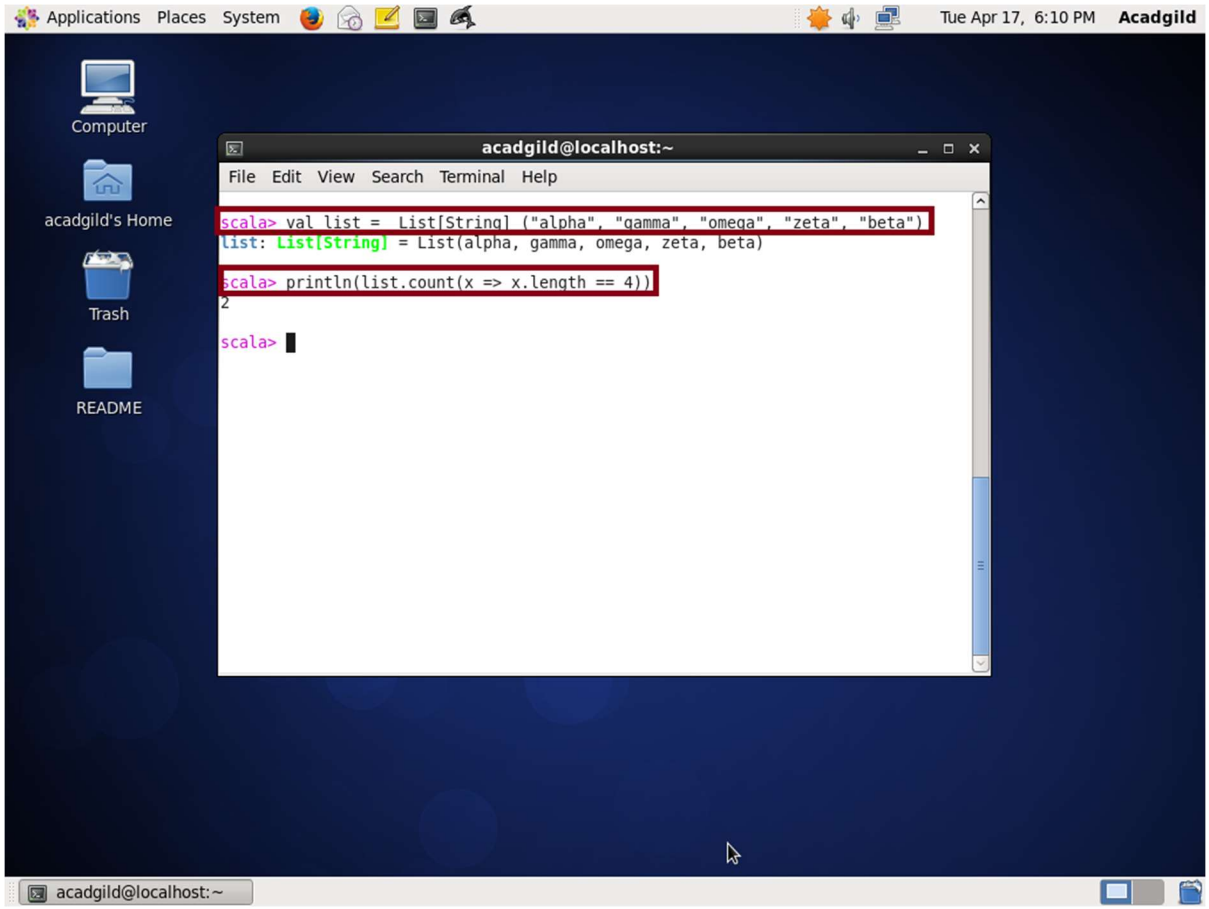
  List of string is declared as

  val list = List[string](“alpha”,”gamma”,”omega”,”zeta”,”beta”)

- **To find the count of all string with length 4**

  By using following command, the count of all strings with length 4 is printed.

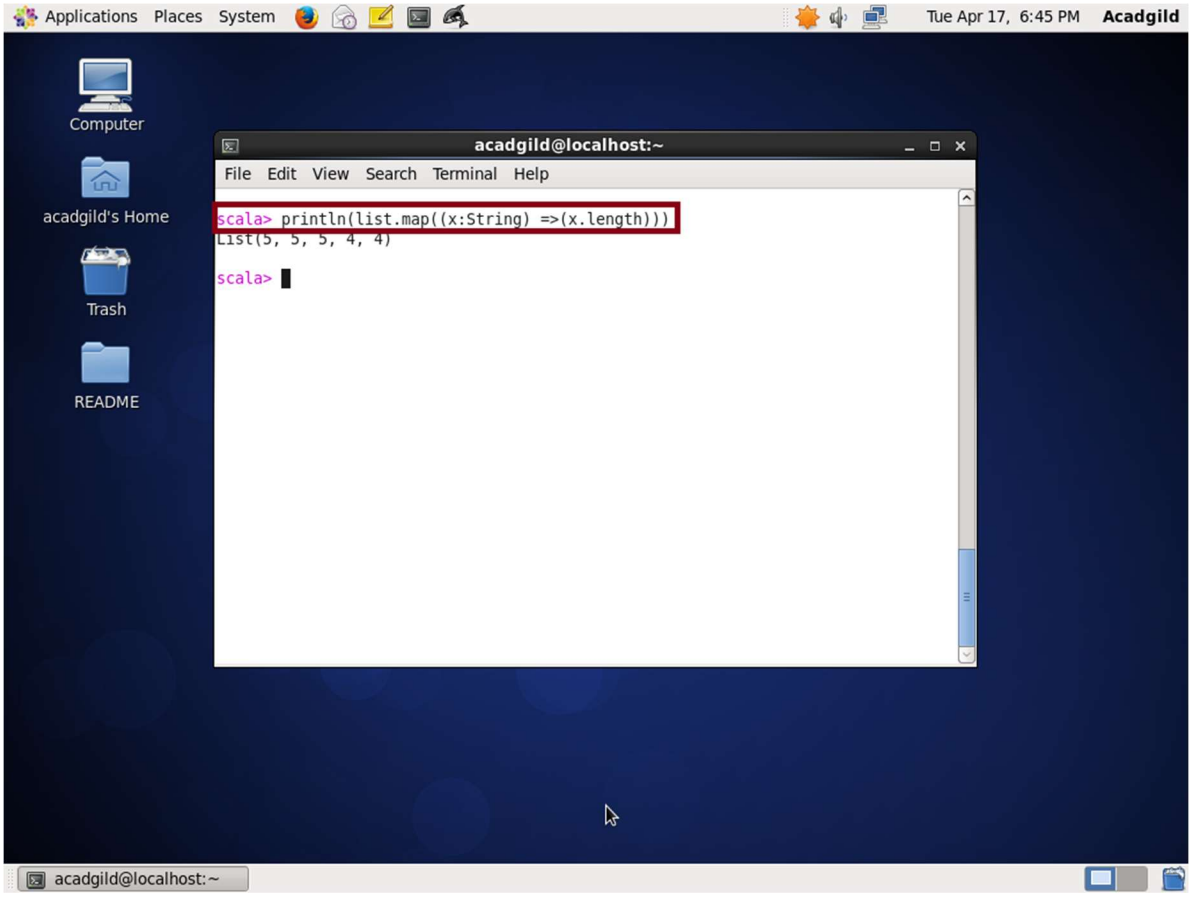
  println(list.count(x=>x.length==4))



- **Convert the list of string to a list of integers, where each string is mapped to its corresponding length.**

  The list of string is converted to a list of integers, where each string is mapped to its corresponding length is printed by the following command

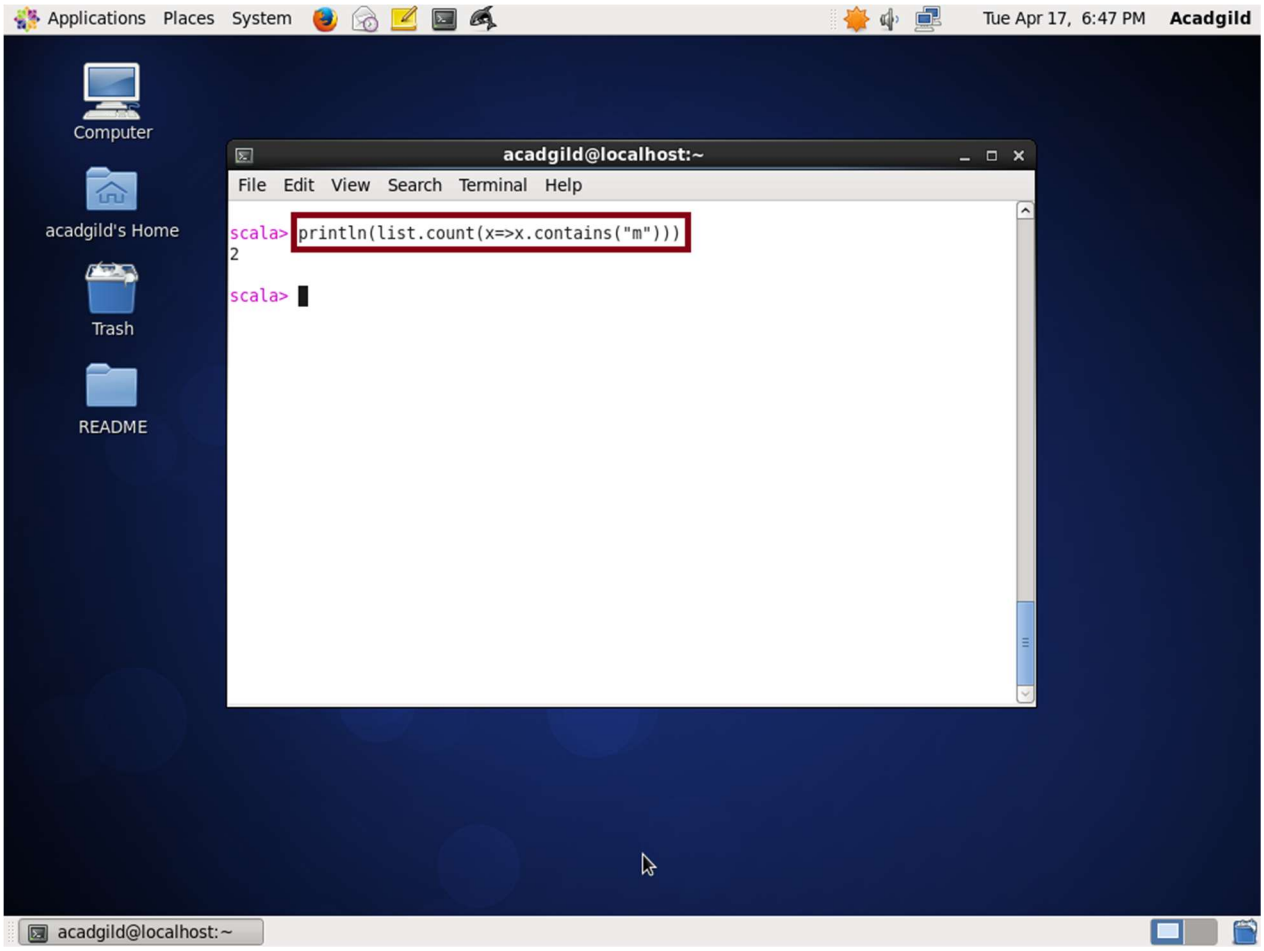
  println(list.map((x:string)=>(x.length)))



➢ **Find count of all strings which contain alphabet ‘m’.**

By using the following command, the count of strings in the list which contains “m” are printed.

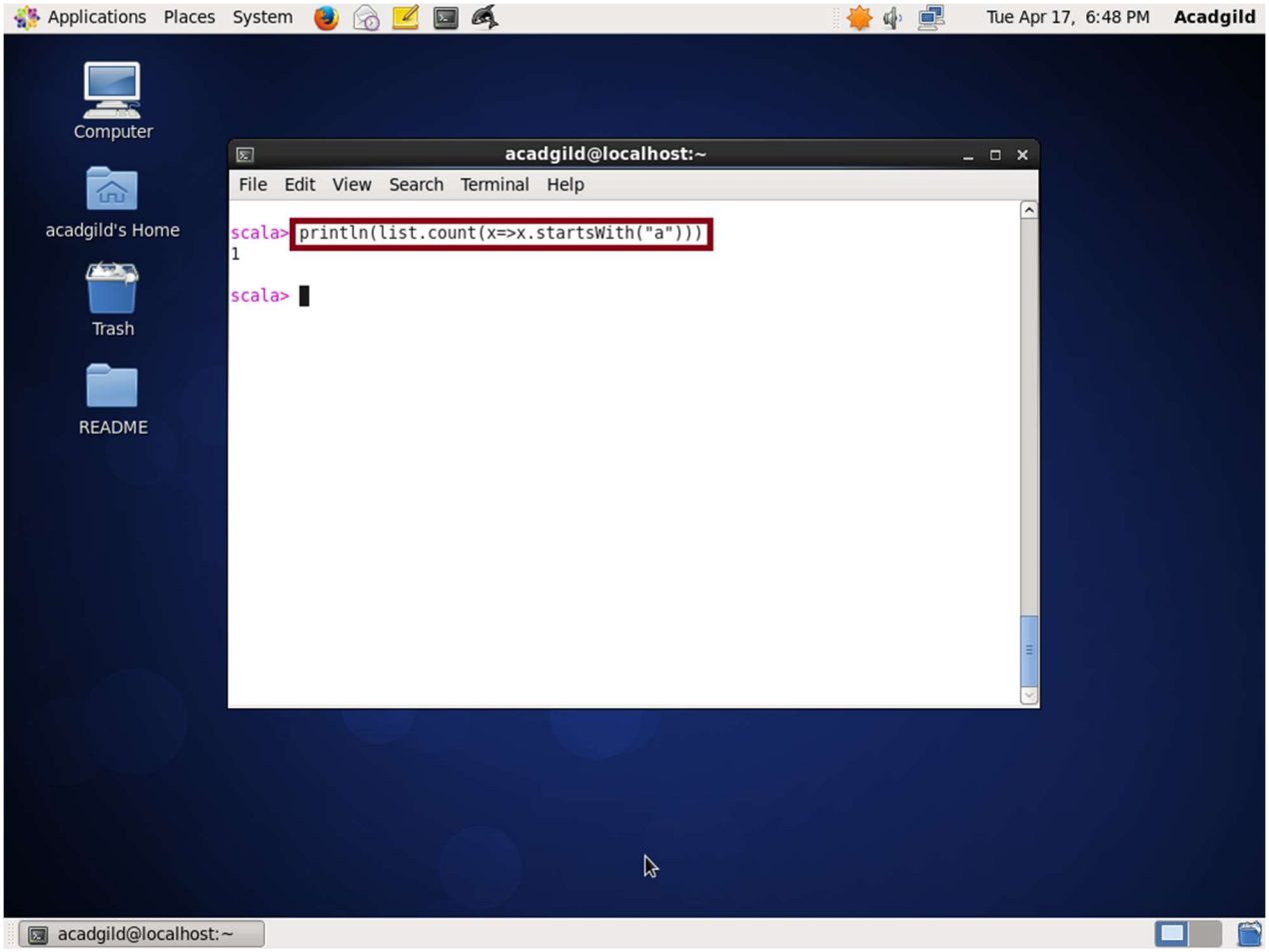
println(list.count(x=>x.contains(“m”)))



- **Find the count of all the strings which start with the alphabet ‘a’.**

  By using the following command, the strings which starts with the alphabet ‘a’ are counted and printed.

  println(list.count(x=>x.startsWith(“a”)))

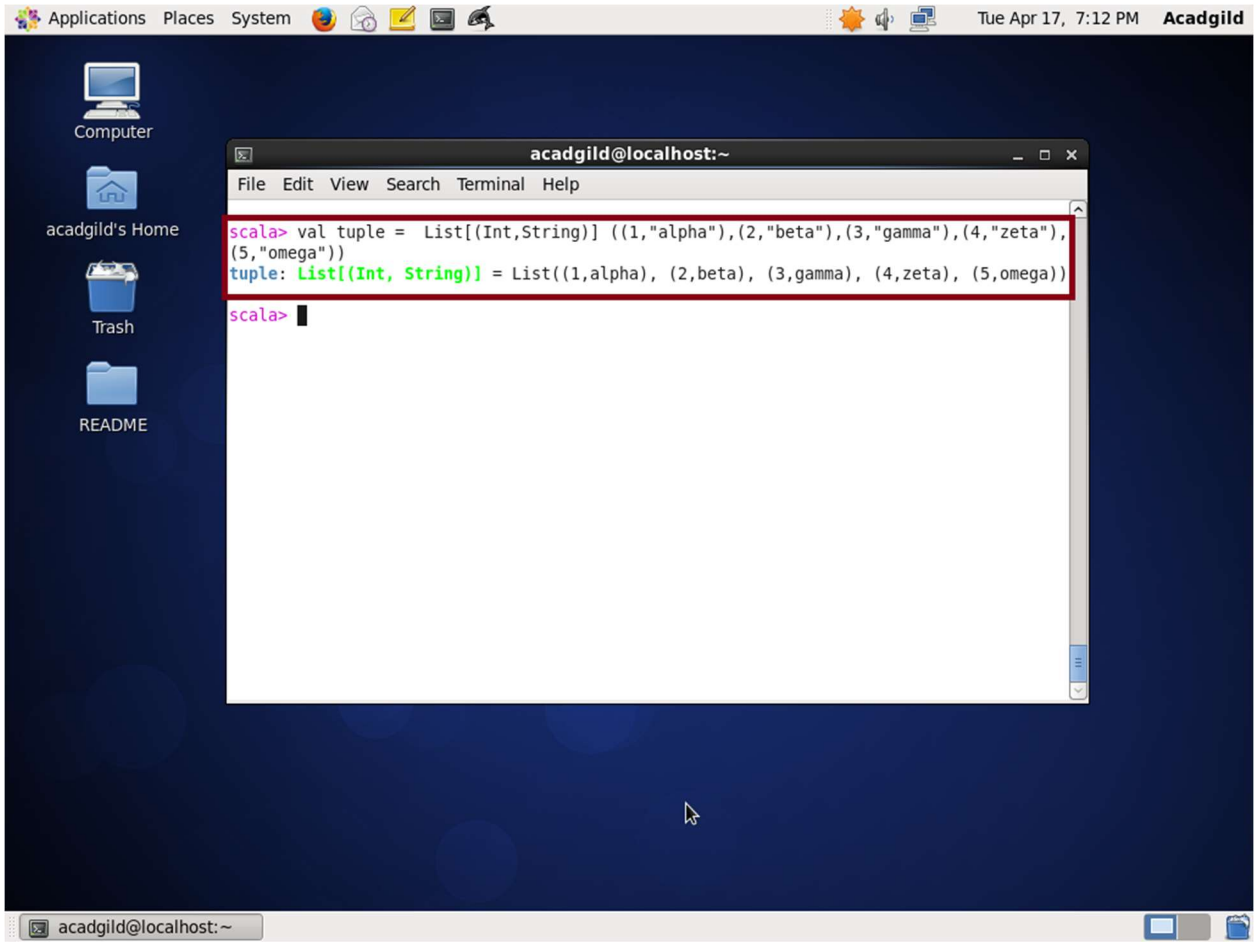


❖ Task 2

➢ **Create a list of tuples, where the 1st element of the tuple is an int and the second element is a string. Example - ((1, 'alpha'), (2, 'beta'), (3, 'gamma'), (4, 'zeta'), (5, 'omega'))**

By using the following command, the list of tuples is created, where the 1st element of the tuple is an int and the second element is a string.

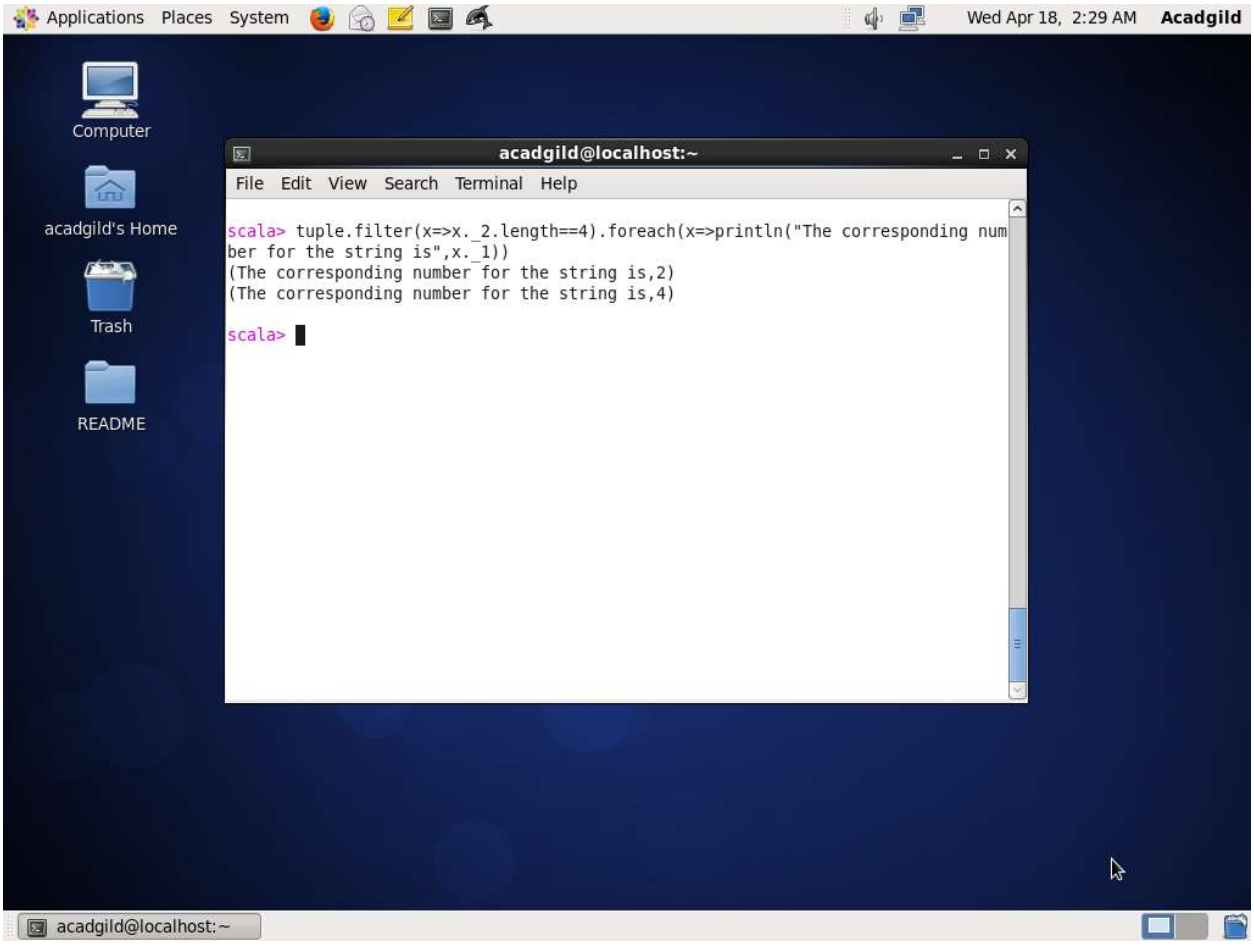
val tuple=List [(Int, String)] ((1," alpha"), (2," beta"), (3," gamma"), (4," zeta"), (5," omega"))



- **For the above list, print the numbers where the corresponding string length is 4.**

By using the following command, the numbers of the corresponding string length is 4 is printed.

tuple.filter(x=>(x._2.length==4)).foreach(x=>println("The corresponding number for the string is",x._1))



- **Find the average of all numbers, where the corresponding string contains alphabet 'm' or alphabet 'z'.**

  By using the following command, the strings containing" m" or "z" alphabets are filtered.

  val tuple_m_z=tuple.filter(x=>(x._2.contains("m")|| (x=>(x._2.contains("z")))

By using the following command, the length of strings containing" m" or "z" alphabets are found.

val tuple_len=tuple_m_z.length

By using the following command, the integers which are mapped to strings containing" m" or "z" alphabets are summed up together.

val tuple_add=tuple_m_z.map(x=>(x._1)). sum

By using the following command, the average of all numbers, where the corresponding string contains alphabet 'm' or alphabet 'z' are determined.

val tuple_avg=tuple_add/tuple_len